Stochastic Processes and their Applications 21 (1985) 107-118
North-Holland

ON THE RESIDUALS OF AUTOREGRESSIVE PROCESSES AND POLYNOMIAL REGRESSION

R.J. KULPERGER
University of Western Ontario, London, Canada

Received 7 February 1984
Revised 1 April 1985

The residual processes of a stationary AR(p) process and of polynomial regression are considered. The residuals are obtained from ordinary least squares fitting. In the AR case, the partial sums converge to Brownian motion. In the polynomial case, they converge to generalized Brownian bridges. Other uses of the residuals are considered. Parameter estimation based on approximate log likelihood function of the residuals is considered.

AMS 1980 Subject Classification: 60F05, 62M10.

auto-regression * polynomial regression * weak convergence * approximate likelihood * least squares * residuals * estimation

1. Introduction

We consider the two regression processes, the stationary autoregressive order p (AR(p)) process (1.1) and polynomial regression (1.2)

$$X_{t+1}=\beta_1X_t+\beta_2X_{t-1}+\cdots+\beta_pX_{t+1-p}+\varepsilon_{t+1}$$

and

$$X_{i,n}=\beta_0+\beta_1t_{i,n}+\cdots+\beta_pt_{i,n}^p+\varepsilon_{i,n}$$

where $\{\varepsilon_t\}$ is an independent and identically distributed (iid) sequence, $E(\varepsilon_t)=0$, $\mathrm{Var}(\varepsilon_t)=\sigma^2$, $E(\varepsilon_t^4)<\infty$ and $\{\varepsilon_{i,n}\}$ is an iid array with $E(\varepsilon_{i,n})=0$, $\mathrm{Var}(\varepsilon_{i,n})=\sigma^2$, $E(\varepsilon_{i,n}^4)<\infty$.

First, consider the autoregressive case. Observe data $X_{-p+1}, X_{-p+2},\ldots,X_n$. The β's are estimated by ordinary least squares, say $\hat{\boldsymbol{\beta}}_n=(\hat\beta_{1,n},\ldots,\hat\beta_{p,n})^t$. The sequence of residuals is defined by

$$\hat\varepsilon_{t,n}=X_{t,n}-(\hat\beta_{1,n}X_{t-1}+\cdots+\hat\beta_{p,n}X_{t-p}),\quad t=1,\ldots,n.$$

$\hat\varepsilon_{1,n},\ldots,\hat\varepsilon_{n,n}$ are not iid, but they are close to $\varepsilon_1,\ldots,\varepsilon_n$, which are iid. As noted in Brown *et al.* (1975), results of residuals of polynomial regression do not generally carry over to autoregressive processes.

Supported by NSERC grant number A5724.

In this note, we consider first partial sums of the residuals of the AR process and show it converges to standard Brownian motion. As another example, we consider an estimation method which seems to be used in practice. Suppose ε_t has density $f(\cdot, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a parameter. By pretending $\hat{\varepsilon}_{1,n}, \ldots, \hat{\varepsilon}_{n,n}$ are iid with this density, the log likelihood is used to estimate $\boldsymbol{\theta}$. Under some suitable conditions, we show this approximate log likelihood method has the same asymptotics as if the true ε_t's were used.

The same problems are considered in the polynomial regression case. The polynomial results are not all new; for example the partial sum process is considered by MacNeill (1978).

The technique used is that of Skorokhod representation (Skorokhod (1956), Breiman (1968, p. 293)). The technique is used by many others, for example Pyke and Shorack (1968). The results are obtained in a straightforward manner by this technique, while they may not be so easy to obtain otherwise. Viewing MacNeill's (1978) polynomial regression results this way makes his work esthetically more appealing.

The results of this paper show that the autoregressive and polynomial residual processes are in some aspects asymptotically similar and in some aspects quite different, for example, partial sums. Finally, an example is given in Section 4 to show that not all functionals of the AR residuals are like those of the corresponding innovations sequence.

2. Asymptotic representation of the AR residuals and some limit theorems

Observe data $X_{-p+1}, X_{-p+2}, \ldots, X_n$. Let $C(k) = \mathrm{Cov}(X_i, X_{i+k})$. Assume that

$$\Sigma = \begin{bmatrix} C(0) & C(1) & \cdots & C(p-1) \\ C(1) & C(0) & \cdots & C(p-2) \\ \vdots & & & \vdots \\ C(p-1) & C(p-2) & \cdots & C(0) \end{bmatrix}$$

is invertible. Let

$$\boldsymbol{X}_n = \begin{bmatrix} X_0 & X_{-1} & \cdots & X_{-p+1} \\ X_1 & X_0 & \cdots & X_{-p+2} \\ \vdots & & & \vdots \\ X_n & X_{n-1} & \cdots & X_{n-p+1} \end{bmatrix}$$

and

$$\boldsymbol{Y}_n^t = (X_1, X_2, \ldots, X_n).$$

For large n, $(1/n)\boldsymbol{X}_n^t \boldsymbol{X}_n$ is invertible, since it converges to Σ. The least squares estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_n = (\boldsymbol{X}_n^t \boldsymbol{X}_n)^{-1} \boldsymbol{X}_n^t \boldsymbol{Y}_n. \tag{2.1}$$

Let $B_t^{(n)} = (1/\sigma\sqrt{n}) \sum_{i=1}^{[nt]} \varepsilon_i$, $0 \le t \le 1$.

Lemma 2.1. *Let $\boldsymbol{\beta}$ be the true value of the parameter. Then $(\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}), B_t^{(n)}, 0 \le t \le 1)$ converges weakly to $(\boldsymbol{Z}, B_t, 0 \le t \le 1)$ where $z^t = (Z_1, \ldots, Z_p)$ is a Gaussian random vector and B_t is a standard Brownian motion.*

Proof. (i) First we show $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ and $B_{t_1}^{(n)}, \ldots, B_{t_k}^{(n)}$ have a joint limiting normal distribution, where $0 \le t_1 \le \cdots \le t_k \le 1$. Since $(1/n)\boldsymbol{X}_n^t \boldsymbol{X}_n \xrightarrow{\Pr} \boldsymbol{\Sigma}$ it suffices to show

$$\frac{1}{n}(\mathbf{X}_n^t \mathbf{X}_n)\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^t = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{i-1}, \ldots, X_{i-p})\varepsilon_i$$

and $B_{t_1}^{(n)}, \ldots, B_{t_n}^{(n)}$ have a joint limiting normal distribution.

Let $a_1, \ldots, a_m \in \mathbb{R}$ be arbitrary, $m = k + p$, and

$$\begin{aligned} Y_n &= a_1 \sum_{i_1=1}^{n} X_{i_1 - j_1}\varepsilon_{i_1} + \cdots + a_k \sum_{i_k=1}^{n} X_{i_k - j_k}\varepsilon_{i_k} \\ &\quad + a_{i_{k+1}} \sum_{i_{k+1}=1}^{n} \varepsilon_{i_{k+1}} + \cdots + a_{i_m} \sum_{i_m=1}^{n} \varepsilon_{i_m} \\ &= \sum_{i=1}^{n} T_i \varepsilon_i \end{aligned}$$

where T_i is measurable with respect to $\mathcal{F}_i = \sigma(\varepsilon_j : j \le i)$. $\{Y_n\}$ is a martingale with respect to $\{\mathcal{F}_n\}$. By a law of large numbers argument $(1/n)\sum_{i=1}^{n} T_i^2 \xrightarrow{\Pr}$ constant, which depends on $a_1, \ldots, a_n$. Thus by the martingale CLT (Hall and Heyde (1980)) $n^{-1/2}Y_n$ has a normal limit. Since $a_1, \ldots, a_m$ are arbitrary, the convergence of finite dimensional distribution is obtained.

(ii) The tightness follows immediately.

$\{\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \beta)$, $B_t^{(n)}$, $0 \le t \le 1$ is an $(\mathbb{R}^p \times D[0,1],\ E^p \times J_1)$-valued process, where E^p is the Euclidean topology on $\mathbb{R}^d$ and J_1 is the Skorkhod topology on $D[0, 1]$. The tightness on this product space will now follow easily from the characterization in Billingsley (1968, Theorem 15.2). □

$D[0, 1]$, the space of right continuous functions with left hand limits, when equipped with the Skorokhod topology, is a complete separable metric space (Billingsley (1968)). By Theorem 3.9.2 of Skorokhod (1956), there exists a probability space on which processes $D_t^{(n)}$, B_t, $0 \le t \le 1$ and random vectors $\boldsymbol{Z}_n$ and $\boldsymbol{Z}$ are constructed such that

(i) $(\boldsymbol{Z}, B_t, 0 \le t \le 1)$ has the limit distribution in Lemma 1, (2.2a)

(ii) $(\boldsymbol{Z}_n, D_t^{(n)}, 0 \le t \le 1) \stackrel{\mathrm{D}}{=} (\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}), B_t^{(n)}, 0 \le t \le 1)$ for each n, (2.2b) and

(iii) since B_t is continuous,

$$\delta_n = \max\left\{|Z_n - Z|, \sup_{0\le t\le 1} |D_t^{(n)} - B_t|\right\} \to 0 \quad \text{a.s.} \tag{2.2c}$$

where $\overset{D}{=}$ means equal in distribution. Skorokhod's result does not say anything about the joint distribution of the processes over various n.

Let $\varepsilon_{i,n}^* = \sigma\sqrt{n}(D_{i/n}^{(n)} - D_{(i-1)/n}^{(n)})$, $i = 1, \ldots, n$. Given $(X_{-p+1,n}, \ldots, X_{0,n}) = (X_{-p+1}, \ldots, X_0)$, define

$$X_{i+1,n} = \beta_1 X_{i,n} + \cdots + \beta_p X_{i-p,n} + \varepsilon_{i+1,n}, \quad i = 0, 1, \ldots, n.$$

Then, for each $n \ge 1$,

$$(\varepsilon_{1,n}^*, \ldots, \varepsilon_{n,n}^*, X_{0,n}, \ldots, X_{n,n}) \overset{D}{=} (\varepsilon_1, \ldots, \varepsilon_n, X_0, \ldots, X_n).$$

Let $\hat{\boldsymbol{\beta}}_n^*$ be the least squares estimate of $\boldsymbol{\beta}$ based on data $X_{i,n}$, $i = -p+1, \ldots, n$. The above gives, for each n,

$$\begin{aligned}(\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta}), \varepsilon_{1,n}^*, \ldots, \varepsilon_{n,n}^*) &\overset{D}{=} (\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}), \varepsilon_1, \ldots, \varepsilon_n) \\ &\overset{D}{=} (\boldsymbol{Z}_n^*, \varepsilon_{1,n}^*, \ldots, \varepsilon_{n,n}^*).\end{aligned} \tag{2.3}$$

The residuals are defined by

$$\hat{\varepsilon}_{i,n} = X_i - \sum_{j=1}^p \hat{\beta}_{j,n} X_{i-j} = \varepsilon_i - \sum_{j=1}^p (\hat{\beta}_{j,n} - \beta_j) X_{i-j}, \quad i = 1, \ldots, n.$$

Therefore, for $n \ge 1$, by (2.3)

$$(\hat{\varepsilon}_{1,n}^*, \ldots, \hat{\varepsilon}_{n,n}^*) \overset{D}{=} (\hat{\varepsilon}_{1,n}, \ldots, \hat{\varepsilon}_{n,n}) \tag{2.4}$$

where $\hat{\varepsilon}_{i,n}^* = \varepsilon_{i,n}^* - \sum_{j=1}^p (Z_{j,n}/\sqrt{n}) X_{i-j,n}$ and where $\boldsymbol{Z}_n$ is given in (2.2). From (2.2c), we have

$$\hat{\varepsilon}_{i,n}^* = \varepsilon_{i,n}^* - \sum_{j=1}^p (Z_j + \delta_{j,n}) \frac{1}{\sqrt{n}} X_{i-j,n}, \quad i = 1, \ldots, n, \tag{2.5}$$

where $|\delta_{j,n}| \le \delta_n$. Because of (2.4), we can now drop the superscript * notation and the subscript n in X (2.5) when no confusion will result.

Lemma 2.2

$$\frac{1}{n} \max_{1\le j\le p} \sup_{0\le t\le 1} \left|\sum_{i=1}^{[nt]} X_{i-j}\right| \to 0 \quad a.s.$$

Proof. Since $E(\varepsilon_i^4) < \infty$, it follows that $E(|(1/n)\sum_{i=1}^{[nt]} X_i|^4) \leq kn^{-2}$, for a constant k. The lemma now follows by a standard application of the Borel–Cantelli Lemma (Breiman (1968)). □

Consider the residual partial sum process

$$\hat{B}_t^{(n)} = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{[nt]} \hat{\varepsilon}_{i,n}, \quad 0 \leq t \leq 1$$

$$= B_t^{(n)} - \frac{1}{\sigma} \sum_{j=1}^{p} (Z_j + \delta_{j,n}) \frac{1}{n} \sum_{i=1}^{[nt]} X_{i-j}. \tag{2.6}$$

By Lemma 2.1, the second term in (2.6) tends to zero uniformly in t.

Theorem 2.1. *$\hat{B}_t^{(n)}$ converges weakly to B_t, standard Brownian motion.*

Theorem 2.2. *Let $\Psi \geq 0$ be a function on $[0, 1]$ such that $\int_0^1 t\Psi(t)\,\mathrm{d}t < \infty$. Let*

$$R\left(\frac{i}{n}\right) = \int_{(2i-1)/n}^{(2i+1)/n} \Psi(s)\,\mathrm{d}s.$$

Then

$$\sum_{i=1}^{n} R\left(\frac{i}{n}\right)(\hat{B}_{i/n}^{(n)})^2 \to \int_0^1 \Psi(t) B_t^2\,\mathrm{d}t$$

in distribution.

Proof. From (2.6),

$$\sum_{i=1}^{n} R\left(\frac{i}{n}\right)(\hat{B}_{i/n}^{(n)})^2 = \sum_{i=1}^{n} R\left(\frac{i}{n}\right)(B_{i/n}^{(n)})^2$$
$$- \frac{2}{\sigma} \sum_{i=1}^{n} R\left(\frac{i}{n}\right) B_{i/n}^{(n)} \sum_{j=1}^{p} (Z_j + \delta_{j,n}) \frac{1}{n} \sum_{k=1}^{i} X_{k-j}$$
$$+ \frac{1}{\sigma^2} \sum_{i=1}^{n} R\left(\frac{i}{n}\right)\left(\sum_{j=1}^{p} (Z_j + \delta_{j,n}) \frac{1}{n} \sum_{k=1}^{i} X_{k-j}\right)^2.$$

The result now follows by Lemma 2.2 and Theorem 2.1. □

For $\Psi = 1$, $P(\int_0^1 B_t^2\,\mathrm{d}t \leq x)$ is partially tabled in MacNeill (1978), Table 2 in the column headed Brownian motion.

3. Polynomial regression residual process

Consider the polynomial regression

$$X_{i,n} = \beta_0 + \beta_1 t_{i,n} + \cdots + \beta_p t_{i,n}^p + \varepsilon_{i,n}$$

where $t_{i,n}=i/n$. The analogue of Lemma 2.1 can easily be obtained. Let

$$B_t^{(n)}=\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{[nt]}\varepsilon_{i,n}.$$

Then

$$\begin{aligned}\hat{B}_t^{(n)}&=\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{[nt]}\hat{\varepsilon}_{i,n}=B_t^{(n)}-\sum_{j=0}^{p}(\hat{\beta}_{j,n}-\beta_j)\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{[nt]}t_{i,n}^j\\&\sim B_t^{(n)}-\sum_{j=0}^{p}\frac{Z_j}{\sigma}\int_0^t s^j\,\mathrm{d}s=B_t^{(n)}-\sum_{j=0}^{p}\frac{Z_j}{\sigma}\frac{t^{j+1}}{j+1}.\end{aligned}\tag{3.1}$$

Therefore

$$B_t^{(n)}\to B_t-\sum_{j=0}^{p}\frac{Z_j}{\sigma}\frac{t^{j+1}}{j+1}.\tag{3.2}$$

However

$$0=\sum_{i=1}^{n}t_{i,n}^j\hat{\varepsilon}_{i,n}.$$

Substituting into (3.1), interchanging order of summation, and letting $n\to\infty$ gives

$$\begin{aligned}0&=B_1-j\int_0^1 s^{j-1}B_s\,\mathrm{d}s-\sum_{k=0}^{p}\frac{Z_k}{\sigma}\int_0^1 s^{k+j}\,\mathrm{d}s\\&=B_1-j\int_0^1 s^{j-1}B_s\,\mathrm{d}s-\sum_{k=0}^{p}\frac{Z_k}{\sigma}\frac{1}{k+j+1},\quad j=0,1,\ldots,p.\end{aligned}$$

Therefore

$$\begin{bmatrix}1 & \frac{1}{2} & \cdots & \dfrac{1}{p+1}\\ \frac{1}{2} & \frac{1}{3} & \cdots & \dfrac{1}{p+2}\\ \vdots & \vdots & & \\ \dfrac{1}{p+1} & \dfrac{1}{p+2} & \cdots & \dfrac{1}{2p+1}\end{bmatrix}\begin{bmatrix}\dfrac{Z_0}{\sigma}\\ \dfrac{Z_1}{\sigma}\\ \vdots\\ \dfrac{Z_p}{\sigma}\end{bmatrix}=\begin{bmatrix}B_1\\ B_1-\displaystyle\int_0^1 B_s\,\mathrm{d}s\\ \vdots\\ B_1-P\displaystyle\int_0^1 s^{p-1}B_s\,\mathrm{d}s\end{bmatrix}.$$

Substituting into (3.2) gives the limit of $B_t^{(n)}$.

If $p=1$, we get

$$\frac{Z_0}{\sigma}=-2B_1+6\int_0^1 B_s\,\mathrm{d}s,\qquad \frac{Z_1}{\sigma}=6B_1-12\int_0^1 B_s\,\mathrm{d}s$$

and

$$\hat{B}_t^{(n)}\to B_t-tB_1+6t(1-t)\left(\tfrac{1}{2}B_1-\int_0^1 B_s\,\mathrm{d}s\right)$$

This agrees with MacNeill's (1978) result obtained by less elementary techniques. For $p=2$ we get

$$\hat{B}_t^{(n)} \to (B_t - tB_1) - 3t(t-1)\left(B_1 - 2\int_0^1 B_s\,\mathrm{d}s\right)$$
$$-5t(t-1)(2t-1)\left(B_1 + 6\int_0^1 B_s\,\mathrm{d}s - 12\int_0^1 sB_s\,\mathrm{d}s\right).$$

The difference between this and the AR process is that the last term in (3.1) does not tend to zero, while the last term in (2.6) converges to zero by a SLLN.

4. Applications and remarks

In this section we consider some applications of the asymptotic representation of the residuals. Sometimes data is prefiltered and then the residuals are modelled. In economic series, it is quite common to fit a trend and then model the residuals.

As a first example we consider an approximate likelihood method by pretending $\hat{\varepsilon}_{1,n}, \ldots, \hat{\varepsilon}_{n,n}$ are iid. Suppose ε_n has density $f(\cdot, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\theta}$, an open subset of $\mathbb{R}^d$. If $\varepsilon_1, \ldots, \varepsilon_n$ were observed, $\boldsymbol{\theta}$ could be estimated by the method of maximum likelihood (mle), obtained by maximizing the log likelihood function

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log f(\varepsilon_t, \boldsymbol{\theta}). \tag{4.1}$$

$\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ could also simultaneously be estimated by maximum likelihood. Suppose instead we estimate $\boldsymbol{\theta}$ by maximizing the approximate log likelihood

$$\hat{L}_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log f(\hat{\varepsilon}_{t,n}, \boldsymbol{\theta}), \tag{4.2}$$

which is (4.1) with $\hat{\varepsilon}_{t,n}$ replacing ε_t. Are the approximate mle and true mle, from (4.2) and (4.1) respectively, close? Under the assumptions below they are. As pointed out by the referee, the suboptimal estimation (4.2) is asymptotically efficient (corollary to Theorem 4.1) which seems to be due to the smoothness assumption 4.1(2) on the density f. This assumption is satisfied by many members of the exponential family in their natural parameterizations (Rao (1973)).

We make some assumptions of f, in the spirit of the type used in Rao (1973) to prove consistency and asymptotic normality of the mle.

Assumption 4.1. (1) Suppose $f(x, \boldsymbol{\theta})$ satisfies Assumptions 1, 2 and 3 of Rao (1973, p. 364).

(2) Suppose, for some $m \geq 1$ and for any $\boldsymbol{\theta} \in \boldsymbol{\theta}$, and for almost all x,

$$\frac{\partial}{\partial \boldsymbol{\theta}} \log f(x + \Delta x, \boldsymbol{\theta}) = \sum_{l=0}^{m} \frac{\partial^{l+1}}{\partial x^l \partial \boldsymbol{\theta}} \log f(x, \boldsymbol{\theta}) \frac{(\Delta x)^l}{l!} + R_{m+1}(x^*, \boldsymbol{\theta}) \frac{(\Delta x)^{m+1}}{(m+1)!}$$

where x^* is between x and $x+\Delta x$ and

$$|R_{m+1}(x^*, \boldsymbol{\theta})| \leq M(\boldsymbol{\theta}) < \infty.$$

(3) $$E\left(\left|\frac{\partial^{l+1}}{\partial x^l \partial \boldsymbol{\theta}} \log f(\varepsilon_1, \boldsymbol{\theta})\right|\right) < \infty, \quad l = 0, 1, \ldots, m.$$

Part 2 of Assumption 4.1 is used in the same spirit as Assumption 3, p. 364 of Rao (1973).

Theorem 4.1. *Let $A \subset \Theta$ be a compact set, with $\boldsymbol{\theta}_0 \in \mathrm{interior}(A)$, where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. Then on the specially constructed probability space of Section* 2,

$$\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial \boldsymbol{\theta}} \hat{L}_n(\boldsymbol{\theta}) - \frac{\partial}{\partial \boldsymbol{\theta}} L_n(\boldsymbol{\theta})\right) \to 0 \quad \textit{a.s. uniformly on } A.$$

Corollary to Theorem 4.1. *Let $\hat{\boldsymbol{\theta}}_n$ and $\hat{\hat{\boldsymbol{\theta}}}_n$ respectively minimize L_n and $\hat{L}_n$ on A.*

(1) *On the specially constructed probability space*

$$\hat{\hat{\boldsymbol{\theta}}}_n - \hat{\boldsymbol{\theta}}_n \to 0 \quad \textit{a.s.}$$

Also since $\hat{\boldsymbol{\theta}}_n$ is $\sqrt{n}$-consistent (Rao (1973)) then back in the original probability space, $\hat{\hat{\boldsymbol{\theta}}}_n$ is $\sqrt{n}$-consistent.

(2) *Let $I = -E_{\boldsymbol{\theta}_0}(\partial^2/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta} \log f(\varepsilon_1, \boldsymbol{\theta}_0))$ be the Fisher information. Then*

$$I\sqrt{n}(\hat{\hat{\boldsymbol{\theta}}} - \boldsymbol{\theta}_0) \xrightarrow{\mathrm{D}} N(0, 1).$$

Notice also $\hat{\hat{\boldsymbol{\theta}}}_n$ is asymptotically independent of $\hat{\boldsymbol{\beta}}_n$.

Proof of the Corollary. (1) Follows directly from Theorem 4.1.

For (2),

$$\frac{1}{\sqrt{n}} \frac{\partial L_n}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}}\left(\frac{\partial L_n}{\partial \boldsymbol{\theta}}(\hat{\hat{\boldsymbol{\theta}}}_n) - \frac{\partial \hat{L}_n}{\partial \boldsymbol{\theta}}(\hat{\hat{\boldsymbol{\theta}}}_n)\right) + \frac{1}{\sqrt{n}} \frac{\partial \hat{L}_n}{\partial \boldsymbol{\theta}}(\hat{\hat{\boldsymbol{\theta}}}_n)$$
$$- \frac{1}{n}\left(\frac{\partial^2 L_n}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}}(\boldsymbol{\theta}_n^*)\right)\sqrt{n}(\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_n^*$ is between $\hat{\hat{\boldsymbol{\theta}}}_n$ and $\boldsymbol{\theta}_0$.

On the right-hand side, the first term tends to zero, the second term equals zero, and

$$\frac{1}{n}\left(\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} L_n(\boldsymbol{\theta}_n^*)\right) \to -I$$

The corollary now follows by Slutsky's Theorem. Also we see that on this specially constructed space, $\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_0$ asymptotically does not depend on $Z_1, \ldots, Z_p$. □

Proof of Theorem 4.1

$$\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial \boldsymbol{\theta}}\hat{L}_n(\boldsymbol{\theta})-\frac{\partial}{\partial \boldsymbol{\theta}}L_n(\boldsymbol{\theta})\right)$$

$$=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\sum_{l=1}^{m}\frac{\partial^{l+1}}{\partial \boldsymbol{\theta}}\log f(\varepsilon_t,\boldsymbol{\theta})\frac{(-1)^l}{l!}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{p}(Z_j+\delta_{jn})X_{t-j}\right)^l\right.$$

$$\left.+R_{m+1,i}\left(-\frac{1}{\sqrt{n}}\sum_{j=1}^{p}(Z_j+\delta_{jn})X_{t-j}\right)^{m+1}\right\}.$$

For $l=1$, we get terms involving

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial^2}{\partial x\partial \boldsymbol{\theta}}\log f(\varepsilon_t,\boldsymbol{\theta})\frac{X_{t-j}}{\sqrt{n}}=\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2}{\partial x\partial \boldsymbol{\theta}}\log f(\varepsilon_t,\boldsymbol{\theta})X_{t-j}$$

$$\to E_{\theta_0}\left(\frac{\partial^2}{\partial x\partial \boldsymbol{\theta}}\log f(\varepsilon_1,\boldsymbol{\theta})X_{1-j}\right)\quad\text{by a SLNN argument}$$

$$=E_{\theta_0}\left(\frac{\partial^2}{\partial x\partial \boldsymbol{\theta}}\log f(\varepsilon_1,\boldsymbol{\theta})\right)E_{\theta_0}(X_{1-j})\quad\text{by independence, since } j=1,\ldots,p$$

$$=0.$$

For $l=2,\ldots,m$, we have terms, involving, for $k=0,1,\ldots,l$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{\partial^{l+1}}{\partial x^l\partial \boldsymbol{\theta}}\log f(\varepsilon_t,\boldsymbol{\theta})(X_{t-j})^k(\sqrt{n})^{-l}$$

$$=\left(\frac{1}{\sqrt{n}}\right)^{(l-1)}n^{-1}\sum_{t=1}^{n}\frac{\partial^{l+1}}{\partial x^l}\log f(\varepsilon_t,\boldsymbol{\theta})X_{t-j}^k$$

$$\to 0\quad\text{by a SLNN argument.}$$

For the remainder term, for each j an upper bound is, for n sufficiently large

$$(|Z_j|+1)\frac{M(\boldsymbol{\theta})}{n^{m-2}}\sum_{t=1}^{n}|X_{t-j}|^k\sim(|Z_j|+1)M(\boldsymbol{\theta})E_{\theta_0}(|X_{1-j}|^k)n^{-m/2}\to 0.$$

In the polynomial case, under Assumption 4.1, we get

Theorem 4.2. *On the specially constructed probability space,*

$$\frac{1}{\sqrt{n}}\left(\frac{\partial}{\partial \boldsymbol{\theta}}\hat{L}_n(\boldsymbol{\theta})-\frac{\partial}{\partial \boldsymbol{\theta}}L_n(\boldsymbol{\theta})\right)\to-\sum_{j=0}^{p}Z_jE_{\theta_0}\left(\frac{\partial^2}{\partial x\partial \boldsymbol{\theta}}\log f(\varepsilon_1,\boldsymbol{\theta})\right)\int_0^1 s^j\,\mathrm{d}s.$$

uniformly on compact sets in $\boldsymbol{\theta}$. Here L_n and $\hat{L}_n$ are the polynomial case analogues of (4.1) *and* (4.2).

Remark. In the polynomial regression problem we see that the approximate log likelihood differs asymptotically from the mle unless

$$E_{\boldsymbol{\theta}_0}\left(\frac{\partial^2}{\partial x \partial \boldsymbol{\theta}} \log f(\varepsilon_1, \boldsymbol{\theta}_0)\right) = 0.$$

We now briefly consider the empirical distribution function (edf) of the autoregressive residuals. On a specially constructed space, as in Section 2, there also exists a standard Brownian bridge W, so that for F_n, the edf of the unobserved innovations $\varepsilon_1, \ldots, \varepsilon_n$,

$$\delta_n(x) = \sqrt{n}(F_n(x) - F(x)) - W(F(x)) \to 0 \tag{4.3}$$

uniformly in x, a.s. as $n \to \infty$.

Using (4.3) and Lemma 2.1,

$$\hat{F}(x) = n^{-1} \sum_{i=1}^{n} I(\varepsilon_{i,n} \leqslant x) \leqslant F_n\left(x + \frac{1}{\sqrt{n}} \sum_{j=1}^{p} (|z_{jn}| + |\delta_{jn}|) \max_{1 \leqslant t \leqslant n} \left|\sum_{l=1}^{p} X_{t-l}\right|\right).$$

An analogous lower bound on $\hat{F}_n$ is also easily obtained. Using these and (4.3),

$$\begin{aligned} \sup_x |\hat{F}_n(x) - F(x)| &\leqslant \sup_x |F_n(x) - F(x)| + \sup_x \sup_s{}^* |F(x+s) - F(x)| \\ &\quad + \frac{1}{\sqrt{n}} \sup_x \sup_s{}^* |W(F(x+s)) - W(F(x))| \\ &\quad + \frac{1}{\sqrt{n}} \sup_x \sup_s{}^* |\delta_n(x+s) - \delta_n(x)| \end{aligned} \tag{4.4}$$

where $\sup_s^*$ is over all

$$|s| \leqslant \frac{1}{\sqrt{n}} \sum_{j=1}^{p} (|z_j| + |\delta_{jn}|) \max_{1 \leqslant t \leqslant n} \left(\left|\sum_{j=1}^{p} X_{t-j}\right|\right).$$

Lemma 4.1. *For the* Ar(p) *process considered in Section* 2, *we have, for any* $\delta > \frac{1}{4}$,

$$n^{-\delta} \max_{1 \leqslant t \leqslant n} |X_t| \to 0 \quad a.s.$$

Proof. Write

$$X_t = \sum_{l=0}^{t} C_{l,t-l} + \sum_{l=1}^{p} a_{l,t} X_{-l},$$

where $(1 - \beta_1 x - \cdots - \beta_p x^p)^{-1} = \sum_{l=0}^{\infty} C_l x^l$ and $a_{1,t}, \ldots, a_{p,t} \to 0$ as $t \to \infty$. Since ε_t are iid, it is easily shown that

$$E\left(\left(\sum_{l=0}^{t} C_l \varepsilon_{t-l}\right)^4\right) \leqslant \mu_4 \sum_{l=0}^{t} C_l^4 + \mu_2^2 \left(\sum_{l=0}^{t} C_l\right)^2 \leqslant K,$$

independent of t. The result now follows from the Borel–Cantelli Lemma. $\square$

Proposition 4.2. *Suppose, for any* $a>0$, $\frac{1}{4}<\delta<\frac{1}{2}$,

$$\sup_x \sup\{|F(x+s)-F(x)|: |s|\le an^{-(1/2-\delta)}\} = O(n^{-(1/2-\delta)}).$$

Then for any α, $0<\alpha<\frac{1}{2}-\delta$,

$$n^{\alpha}|F_n(x)-F(x)| \xrightarrow{\Pr} 0 \quad \textit{uniformly in } x.$$

Proof. On the specially constructed space, the proposition follows directly from Lemma 4.1 and (4.4). In fact using more moments on ε_i in Lemma 4.1 would allow δ to be decreased in this result. □

Remark. We see that in some respects the AR and polynomial residuals behave similarly and sometimes they behave differently. This elaborates a passing remark in Brown et al. (1975).

The AR(p) residual partial sum process $\hat{B}_t^{(n)}$ of Section 2 can be used in an analogous manner to MacNeill (1978). For example to test H_0: $\boldsymbol{\beta}$ does not change over time, versus H_A: $\boldsymbol{\beta}$ does change over time, a test statistic is

$$Q_n = \sum_{i=1}^{n} R\left(\frac{i}{n}\right)\left(\frac{1}{\alpha\sqrt{n}}\sum_{j=1}^{i}\hat{\varepsilon}_{j,n}\right)^2$$

with a consistent estimate of σ^2 substituted, for example $s^2 = 1/n\sum_{i=1}^{n}\hat{\varepsilon}_{i,n}^2$. Theorem 2.2 gives the asymptotic distribution of this Cramer–von Mises statistic under H_0. The hypothesis H_0 is rejected for large values of Q_n.

A goodness of fit test could be based on

$$S_n = \max_{1\le i\le n}\frac{1}{\sigma\sqrt{n}}\left|\sum_{j=1}^{n}\hat{\varepsilon}_{j,n}\right| = \sup_{0\le t\le 1}|\hat{B}_t^{(n)}|.$$

If X_t is an AR(p) process, then Theorem 2.1 gives the asymptotic distribution of S_n. A variety of tests based on the residual edf could also be considered.

Not all functions of autoregressive residuals have the same asymptotics as the corresponding function of the innovation sequence. We end this paper with one such example. A statistic of interest in time series is the spectral measure (Brillinger (1975)). Bartlett (1962) considered such a statistic. Suppose one observes a stationary AR(1) process

$$X_{n+1} = \beta X_n + \varepsilon_{n+1} \tag{4.5}$$

Finite Fourier transforms of the residuals may be formed,

$$\hat{d}_n(\lambda) = \sum_{j=1}^{n} e^{-i\lambda j}\hat{\varepsilon}_{j,n}.$$

$\hat{I}_n(\lambda)=\hat{d}_n(\lambda)\hat{d}_n(-\lambda)/(2\pi n)$ is the periodogram of the residual sequence. Let

$$\hat{F}_{\varepsilon,n}(\lambda)=\frac{2\pi}{n}\sum_{j=1}^{[n\lambda/2\pi]}\hat{I}_n\left(\frac{2\pi j}{n}\right), \tag{4.6}$$

$F_{\varepsilon,n}$ is an estimate of the empirical spectral measure of the iid sequence $\{\varepsilon_n\}$. Let $F_{\varepsilon,n}$ be the corresponding functional of (4.6) for the $\{\varepsilon_n\}$ sequence. Suppose the true value of β in (4.5) is 0. In this case it can be shown that

$$\begin{aligned}\sqrt{n}\{\hat{F}_{\varepsilon,n}(\lambda)-F_{\varepsilon,n}(\lambda)\}&=\sqrt{n}(\beta-\hat{\beta}_n)\frac{\sigma^2}{\pi}\int_0^\lambda\cos(\gamma)\,\mathrm{d}\gamma+\mathrm{O}_p(1)\\&\xrightarrow{\mathrm{D}}N\left(0,\frac{\sigma^4}{\pi^2}\left(\int_0^\lambda\cos(\gamma)\,\mathrm{d}\gamma\right)^2\right)\end{aligned}$$

since $f(\gamma)=\sigma^2/2\pi$ is the spectral density of $\{\varepsilon_n\}$. At the rate of interest, namely $\sqrt{n}$, there is a non-trivial difference.

References

M.S. Bartlett, An Introduction to Stochastic Processes (Cambridge University Press, London, 1962).

P. Billingsley, Convergence of Probability Measures (J. Wiley and Sons, New York, 1968).

L. Breiman, L., Probability (Addison-Wesley, Reading, MA, 1968).

D.R. Brillinger, Time Series: Data Analysis and Theory (Holt, Rinehart and Winston, New York, 1975).

R.L. Brown, J. Durbin and J.M. Evans, Techniques for testing the constancy of regression relationships over time (with discussion), JRSS B 37 (1975) 149-192.

P. Hall and C.C. Heyde, Martingale Limit Theory and its Application (Academic Press, New York, 1980).

I. B. MacNeill, Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times, Ann. Statist. 6 (1978) 422-433.

R. Pyke and G. Shorack, Weak convergence of a two sample empirical process and a new approach to Chernoff-Savage theorems, Ann. Math. Statist. 39 (1968) 755-771.

C.R. Rao, Linear Statistical Inference and its Applications, second edition (John Wiley and Sons, New York, 1973).

A.V. Skorokhod, Limit Theorems for stochastic processes, Th. Prob. Appl. 1 (1956) 261-290.